Vision Research 51 (2011) 969–977

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

The application of 3D representations in face recognition

Adrian Schwaninger [a,b], Jisien Yang [c,d,*]

[a] School of Applied Psychology, University of Applied Sciences Northwestern Switzerland
[b] Department of Informatics, University of Zurich, Switzerland
[c] Department of Psychology, University of Zurich, Switzerland
[d] Department of Psychology, National Chung Cheng University, Taiwan

ARTICLE INFO

Article history:
Received 14 May 2010
Received in revised form 5 February 2011
Available online 13 February 2011

Keywords:
Silhouette
Face recognition
3D face representation

ABSTRACT

Most current psychological theories of face recognition suggest that faces are stored as multiple 2D views. This research aims to explore the application of 3D face representations by means of a new paradigm. Participants were required to match frontal views of faces to silhouettes of the same faces. The formats of the face stimuli were modified in different experiments to make 3D representations accessible (Experiments 1 and 2) or inaccessible (Experiment 3). Multiple 2D view-based algorithms were not applicable due to the singularity of the frontal-view faces. The results disclosed the application and adaptability of 3D face representations. Participants can readily solve the tasks when the face images retain the information essential for the formation of a 3D face representations. However, the performance substantially declined when the 3D information in faces was eliminated (Experiment 3). Performance also varied between different face orientations and different participant groups.

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction

Face recognition plays an important role in social interaction. Despite all faces being composed of relatively few parts and sharing a similar configuration, humans are able to distinguish exceedingly subtle differences between them. Although much research has been conducted in recent decades to investigate the underlying processes and representations of faces (for a recent review see for example Schwaninger, Wallraven, Cunningham, & Chiller-Glaus, 2006), it remains debatable whether 3D representations and viewpoint transformations exist in human face recognition.

According to object-centered theories (Biederman, 1987, 2000; Marr & Nishihara, 1978), object recognition is based on structural descriptions which specify an object by its constituent parts, e.g. Geons and their spatial relations (Geon Structural Descriptions, GSD). Such descriptions are assumed to be object-centered, which provide the basis for view-invariant recognition. Biederman and Gerhardstein (1993) have enhanced the Recognition by Components (RBC) theory (Biederman, 1987) by specifying three prerequisites for viewpoint-independent recognition. First, the objects must be decomposable into their parts. Second, the GSD for different objects must be distinctive. Third, the same GSD of a specific object must be recoverable from different viewpoints.

In contrast, view-based theories propose that objects are not stored as object-centered structural descriptions but as a collection of 2D views (Biederman & Kalocsai, 1997; Bülthoff & Edelman, 1992; Bülthoff, Edelman, & Tarr, 1995; Tarr & Bülthoff, 1995; Tarr & Pinker, 1989). Object recognition relies on matching a novel view of an object to the stored views by using different mechanisms, such as linear interpolation between views (Poggio & Edelman, 1990), multiple views plus transformations (Tarr & Pinker, 1989), or linear combination of views (Ullman & Basri, 1991).

However, face recognition, it has been suggested, implicates mechanisms that are different from those applying to object recognition. The processes which distinguish object recognition from face recognition have been demonstrated in different research domains, such as behavioral studies (Yin, 1969), neuropsychological patients (Ellis & Florence, 1990; Farah, 1991; Farah, Levinson, & Klein, 1995; Hecaen & Angelergues, 1962; Yin, 1970), and cognitive neuroscience studies (Desimone, 1991; Kanwisher, Downing, Epstein, & Kourtzi, 2001; Kanwisher, McDermott, & Chun, 1997; Ojemann, Ojemann, & Lettich, 1992). However, in contrast to the debate on 2D vs. 3D representation in object recognition, it is generally assumed that faces are represented by a collection of 2D views (Biederman & Kalocsai, 1997; Bülthoff et al., 1995).

Although Biederman and Kalocsai (1997) propose that object recognition is viewpoint-independent, they further indicate that RBC applies only to basic-level object recognition but not to face recognition. Due to the fact that all faces share the same basic components (eyes, nose, mouth, chin, etc.) in the same basic arrangement (the eyes are above the nose which is above the mouth),

* Corresponding author at: Department of Psychology, University of Zurich, Thurgauerstrasse 39, CH-8050 Zurich, Switzerland. Fax: +41 43 336 01 00.
E-mail address: yangjisin@yahoo.com.tw (J. Yang).

0042-6989/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2011.02.006

faces cannot be distinguished based on structural descriptions. Violation of the prerequisites of viewpoint-independent recognition makes it difficult for face recognition to have 3D representations. Instead, Biederman and Kalocsai (1997) argue that a holistic, viewpoint-dependent system such as the models proposed by Christoph von der Malsburg and his colleagues (Lades et al., 1993; Wiskott, Fellous, Krüger, & von der Malsburg, 1997) explain human face recognition much more aptly. Moreover, several studies also suggest that face recognition is based on purely holistic and view-based processes using 2D representations rather than 3D representations (e.g., Lades et al., 1993; Tanaka & Farah, 1991, 1993; Wiskott et al., 1997). Using a computational model, Wallraven, Schwaninger, and Bülthoff (2005) implemented a view-based approach in which facial features and their spatial relations are stored in separate 2D views, which are temporally associated. As shown in a group of different studies, this model could explain various aspects of human face recognition such as processing component and configural information (Schwaninger, Wallraven, & Bülthoff, 2004; Schwaninger, Lobmaier, Wallraven, & Collishaw, 2009) as well as specific effects of viewpoint (Schwaninger, Schumacher, Wallraven, & Bülthoff, 2007; Wallraven, Schwaninger, Schuhmacher, & Bülthoff, 2002).

Accordingly, 3D face representations, which require the processes of utilizing the shading and shadow information to reconstruct the three-dimensional shape of faces, may be absent for humans in many face recognition models (Bruce, 1988; Bruce & Langton, 1994; Johnson, Hill, & Carman, 1992; Vetter, 1998). It is only in the domain of computer vision that 3D models and processes of faces have been implemented (Blanz & Vetter, 1999; Vetter, 1998; for a review of cognitive and computational models of face recognition see Schwaninger et al., 2006).

However, at least three different research lines provide converging results suggesting that human face recognition may involve 3D representation mechanisms rather than a mere match of multiple 2D face views. The first line of evidence comes from research regarding the recognition of one's own profile. Troje and Kersten (1999) have found that humans can recognize profile views of their own faces, even though such views are usually not encountered and thus are hardly available in visual memory. Tong and Nakayama (1999) report further interesting research. In a visual search task, their participants demonstrated an equivalent own-face advantage across frontal, three-quarter, and profile views. Their results are surprising when taking into account that the observers had equal amounts of visual experience of the stranger's frontal and profile view, but far greater experience of their own frontal face view than their profiles. They argue that people can develop robust representations for highly over-learned faces, such as one's own. This representation might involve viewpoint-independent 3D representations. However, participants in Tong and Nakayama's study might have relied on facial texture information to recognize the depth-rotated profiles of their own faces because it has been found that facial information, such as skin color, pigment, or texture features provide information which is important in reducing viewpoint dependence in face recognition (Hill, Schyns, & Akamatsu, 1997). Moreover, face features such as skin texture, blemishes, and dimples may be visible from largely different viewpoints (ÓToole, Bülthoff, Troje, & &Vetter, 1995).

The second line of evidence suggesting the application of 3D face representations in humans comes from haptic face recognition research. Kilgour and Lederman (2002) found that participants' performance (67.8%) was higher than chance (33.3%) when they were shown motionless live faces, and subsequently were required to recognize them by touching (haptic recognition). Their results imply that participants can construct a multimodal 3D representation of a face based on mere visual exposure. Casey and Newell's (2003) research regarding haptic own-face recognition also supports this assumption. They found that a greater amount of target faces were correctly identified when own-face masks were oriented towards participants, contrary to the orientation in which a haptic representation of one's own face is naturally generated. These findings suggest that humans might be able to construct 3D representations of their own faces via the large amount of visual experience. Viewers might then apply this 3D representation to the haptic recognition of representations of their own faces in which geometric properties are the only cues for correct identification. Although far from the same domain, Casey and Newell's results surprisingly correspond to the notion of Tong and Nakayama (1999) that a robust viewpoint-independent representation is formed for highly over-learned faces.

The third line of evidence comes from the studies associated with neuronal activation in the brain. Grill-Spector et al. (1999) found that both the caudal-dorsal (LO) region and the posterior fusiform (PF-LOa) region in the lateral occipital complex (LOC) are maximally activated by images of different individuals' faces. However, the activation is adapted by repeated presentation of identical individual faces, either in the original viewing condition or in a depth rotation condition. The adaptations for the original viewing condition and the depth rotation condition are equivalent in the PF-LOa region. Grill-Spector et al. argue that the PF-Loa is more invariant to changes in the object's position in the visual field compared to LO. Similarly, Chen, Kao, and Tyler (2006) also observed that brain activation is significantly different between frontal-view and inverted faces, but not between frontal-view and 3/4-view faces. These results suggest that there may be neural circuits responsible for the viewpoint invariance of face representation. In fact, a small portion of cells in the macaque superior temporal sulcus (STS) has been observed to respond equally to multiple views of a face (Perrett et al., 1991).

Although the results from the three lines of research imply the application of 3D face representations in humans, these studies did not directly examine the mechanisms of 3D face representations. In this research, we adopt the 'face silhouette vs. frontal-view faces matching' paradigm, in which a one-tone black silhouette is matched with a frontal-view face (Davidenko, 2007). This task can be solved by extracting 3D information, such as shading and shadow information contained in the face photographs (Bruce & Langton, 1994), reconstructing the 3D structure of faces (Bruce et al., 1991; Vetter, 1998), mentally rotating a 3D face model (Blanz & Vetter, 1999; Vetter, 1998), and matching it to the one-tone black silhouette. This task cannot be solved by matching 2D information contained in the faces, such as face configuration, texture, color, blemishes and dimples. Moreover, it cannot be solved by linear combination, because a set of 2D views is necessary for constructing 3D models in this way (Poggio, 1990; Ullman & Basri, 1991). Although, Poggio and his colleagues propose that only one non-accidental 2D view is sufficient for recognition in the case of bilaterally symmetrical objects such as faces, (Beymer & Poggio, 1995; Poggio, 1991; Poggio & Vetter, 1992), they constrain their conclusion by specifying that

> one should avoid to use in the data base a model view which is a fixed point of the symmetry transformations (since the transformation of it generates an identical new view). In the case of faces, this implies that the model view in the data base should not be an exactly front-view (Poggio & Vetter, 1992, p. 15).

Basically, the frontal-view face is singular, and a second view cannot be computed from it (Schyns & Bulthoff, 1993). As a result, a silhouette cannot be generated from just one frontal-view face image by algorithms of different 2D view-based models. According to view-based face recognition theories, matching a frontal-view face to its one-tone black silhouette would be improbable, because both surface information and multiple views integration are not applicable. Moreover, the shape information in the silhouettes is not directly observable in their frontal-view counterparts (Davidenko, 2007). Consequently, a frontal-view face image and its corresponding one-tone black silhouette serve as an optimal pair to examine the application of 3D face representations.

This research aims to adopt this new paradigm to examine whether 3D face representation mechanisms exist and whether they demonstrate typical characteristics found for face recognition such as inversion effect (Yin, 1969, for a review see Valentine, 1988) and own-race advantage (for a review see Meissner & Brigham, 2001). Three experiments were conducted to address this issue. Throughout the research, both Asian and European participants were recruited to compare the potential race effect in face recognition (Goldstein & Chance, 1979; Hayward, Rhodes, & Schwaninger, 2008).

2. Experiment 1

In Experiment 1, we examine whether participants can match frontal-view faces to their corresponding profiles. This experiment was designed to explore how well humans can perform the viewpoint-transformation task under the condition that the face images are different while the texture and pictorial information are retained. The intermediate views between the frontal-view and the profile view were not available. Although the direct pictorial matching of faces was prevented, participants could still rely on other image-based information, e.g. the facial texture, the size and shape of the features, to accomplish the task. The choice of grayscale profiles of the frontal-view faces was based on findings that facial color or reflectance cues play an important role in face recognition (Alley & Schultheis, 2001; Russell, Biederman, Nederhouser, & Sinha, 2007; Yip & Sinha, 2002). Moreover, Hill et al. (1997) also find that facial information, such as skin color, or pigmented or textured features provide information which is important in reducing viewpoint dependence in face recognition. Accordingly, adopting full color images for both frontal-view and profile face images might have encouraged participants merely to rely on skin tone to match the faces. By contrast, Russell et al. (2007) reported that it is more difficult to extract reflectance information for face recognition in grayscale face images. Consequently, we adopted the recognition between full-color frontal-view face and grayscale profiles to reduce the possibility of participants relying only on the reflectance (color) cues during the matching task. The target face and the testing faces were presented sequentially to avoid concurrent pictorial, picture-based recognition.

2.1. Method

2.1.1. Participants

Throughout the whole study (Experiment 1 to Experiment 3), data were collected in both Switzerland and Taiwan. All the participants in Switzerland were Europeans and were recruited from the University of Zurich; all the participants in Taiwan were Asians and were recruited from the National Chung-Cheng University. All of them participated in only one of the three different experiments. Eighteen European students (5 male, mean age = 23 years) from Switzerland and twenty students (11 male, mean age = 24.8 years) from Taiwan participated in Experiment 1.

2.1.2. Materials

Full-color frontal and profile view photographs of 16 Europeans and 16 Asians were taken. Half of these faces were male for each race. The profile images were converted to grayscale images by PhotoImpact 10. To prevent participants from matching the faces by hairstyles and forehead fringes, the distances between the lowest hair cue in the forehead and the concave of the nose of all the faces were measured. The minimum distance among the faces in the same set was taken as the standard length for all the faces in the same set (same face race and gender). The faces in the same set were trimmed to the same extent based on the standard length. In addition, the hair cue that remained visible after the primary trim was further cropped. Fig. 1 displays examples of the front-view faces with their corresponding grayscale profiles.

2.1.3. Procedure

Experiments 1–3 were all three-factor mixed designs with face race and orientation as within-participant factors, and participant group (European vs. Asian) as between-participant factor. All possible combination pairs between the eight exemplars faces in the same set were included, which generated 28 pairs for each face set. Among the 28 pairs for each set, the same frontal-view faces were balanced to show up in the left side or right side of the profile. In addition, the direction (towards right or left) of the profiles and the position of the correct target faces (towards right or left) were also balanced. In half of the trials, the profile faced towards the right and in the other half towards the left. Among the trials of the profiles facing towards the left, half of the correct target frontal-view faces were arranged on the left side (congruous direction) and half on the right sides (incongruous direction). The manipulation was applied also for the trials of the profiles facing towards the right. The 224 trials, 4 face sets $\times$ 28 combinations $\times$ 2 orientations (upright and inverted), were randomly presented in the formal experiment.

In each trial, a frontal-view face was presented on the screen for 3000 ms, which is followed by a blank (black) mask for 500 ms. Two grayscale profiles, one in the left and one in the right, succeeded the mask image. Participants were instructed to judge which of the two profiles was the transformation of the preceding frontal-view face. They were instructed to match either the right or the left grayscale profile to the frontal-view face by pressing the corresponding key. The frontal-view face was consistently presented upright, whereas the grayscale profiles could be presented upright or inverted. The experiment is self-paced without constraint on the presentation time for the response images (two profiles). Participants' heads were not fixed and the viewing distance

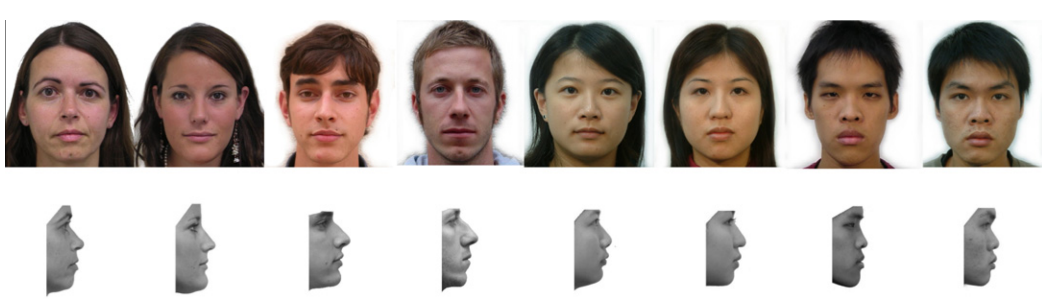

Fig. 1. Examples of the front-view faces with their corresponding grayscale profiles in Experiment 1.

was about 50 cm. Participants were required not to rotate their heads when viewing rotated faces. All the participants were supervised by the same experimenter and were restrained if they rotated their heads during the experiment. Participants were provided with eight practice trials balanced for face race, face direction, orientation, and face gender before the formal experiment. The face images that appeared in the practice trials were not shown again in the formal trials. Fig. 2 illustrates one example of the procedure and the time course of the stimulus presentation in Experiment 1. The rectangles in Fig. 2 denote the border of the 17″ monitor. As can be seen from Fig. 2, the frontal-view face and the two grayscale profiles were not positioned on the same horizontal line so as to reduce direct comparison of the inter-distance between faces' features.

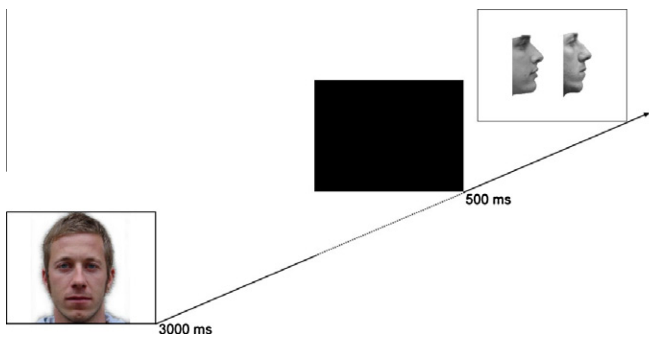

Fig. 2. The procedure and the time course of the stimulus presentation in Experiment 1.

2.2. Results

2.2.1. Accuracy

Throughout the whole research, the data of accuracy were subjected to a repeated measures analysis of variance (ANOVA) with face race and orientation as within-participant factors, and participant group (European vs. Asian) as between-participant factor. Fig. 3 displays means and standard errors of percent correct responses (accuracy) in Experiment 1.

The repeated measures ANOVA revealed main effects of face race, $F(1,36) = 30.99$, $MSE = .13$, $p < .001$, and orientation, $F(1,36) = 122.92$, $MSE = .50$, $p < .001$. The ANOVA also revealed significant interaction between face race and participant group, $F(1,36) = 19.38$, $MSE = .08$, $p < .001$. Simple main effect analysis for the interaction between face race and participant group revealed that European participants performed better in the recognition of European faces ($M = .86$) than matching Asian faces ($M = .76$), $F(1,17) = 45.79$, $MSE = .19$, $p < .001$; but Asian participants performed equally well in the recognition of the two different ethnic faces (European face, $M = .83$; Asian face, $M = .81$), $F < 1$.

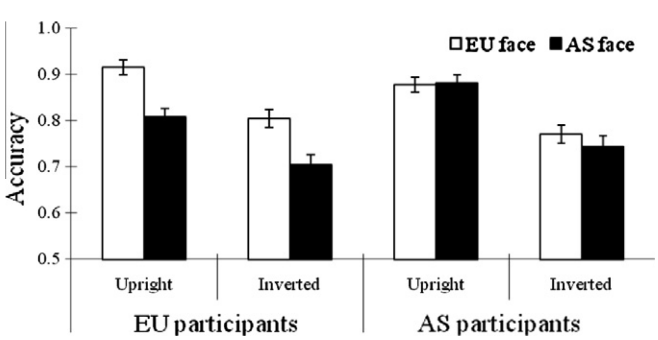

Fig. 3. Means and standard errors of accuracy in different conditions of Experiment 1.

2.2.2. Chance level t-tests

Throughout the whole study, two-tailed *t*-tests were conducted separately for the eight different conditions to test whether or not the performances were higher than chance level (0.5). The performances in all the eight conditions in Experiment 1 were significantly higher than chance level ($ps < .001$).

2.3. Discussion

The results show that participants can readily match the frontal-view faces to their corresponding grayscale profiles. Although the texture information would have been helpful during the task, it would not have been the only information that participants relied on to accomplish the task. The matching process may also involve 3D representations of faces, based on the following inferences. If participants relied merely on the texture information to match the faces, the inversion effect and the asymmetrical race effect would not have been demonstrated. After all, both upright and inverted conditions provide equivalent texture information, and participants from both ethnic groups were provided with the same amount of texture information. The inversion effect and the asymmetrical race effect could possibly have arisen from the distinct 3D face representations between different orientations and between participants from different ethnic groups.

Intriguingly, Experiment 1 demonstrated an asymmetrical race effect for participants from different ethnic groups. European participants performed better with European faces than Asian faces, whereas Asian participants performed comparably well between European and Asian faces. The asymmetric race effect was in accordance with previous research as European participants often reveal race effect, whereas Asian or Africa American participants do not (Barden, Maddux, Petty, & Brewer, 2004; Tanaka, Kiefer, & Bukach, 2004).

3. Experiment 2

In Experiment 2, the grayscale profiles were replaced by one-tone black silhouettes. As discussed earlier, a one-tone black silhouette can prevent participants from relying on any pictorial cue that is available in the profile. This task demanded that participants extract the 3D information from the facial front-view, form a 3D face representation, mentally rotate it, and then match it to the silhouette. The results can provide a direct and critical examination of whether or not humans use 3D representation in face processing.

3.1. Method

3.1.1. Participants

Eighteen European students (1 male, mean age = 24.61 years) from Switzerland and sixteen students (5 male, mean age = 21.52 years) from Taiwan participated in Experiment 2.

3.1.2. Materials

The face stimuli were identical to those used in Experiment 1 but the grayscale profiles were transformed into one-tone black silhouettes by PhotoImpact 10. Fig. 4 displays examples of the front-view faces with corresponding one-tone black silhouettes used in Experiment 2.

3.1.3. Procedure

The design and arrangement of the trials were identical to those in Experiment 1 except that the profiles were replaced by silhouettes. Fig. 5 illustrates the procedure and the time course of the stimulus presentation in Experiment 2.

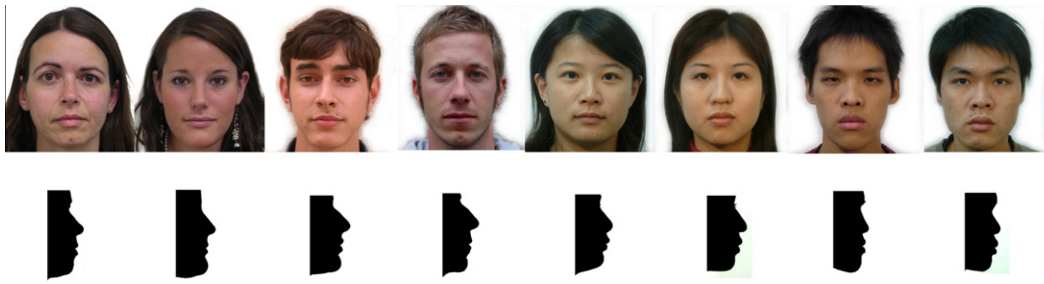

Fig. 4. Examples of the front-view faces with their corresponding one-tone black silhouettes in Experiment 2.

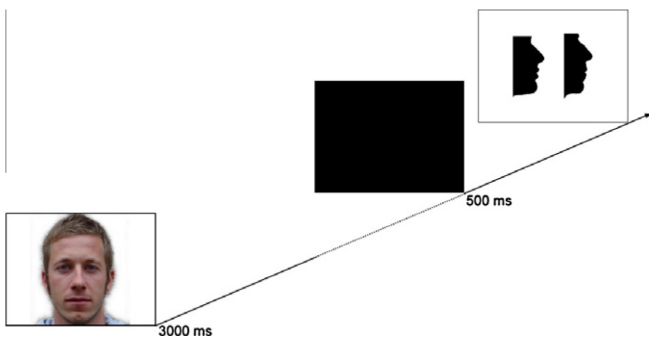

Fig. 5. The procedure and the time course of the stimulus presentation in Experiment 2.

3.2. Results

3.2.1. Accuracy

Fig. 6 displays means and standard errors of accuracy in Experiment 2. The repeated measures ANOVA revealed main effects of face race, $F(1, 32) = 25.76$, $MSE = .08$, $p < .001$, and orientation, $F(1, 32) = 57.50$, $MSE = .20$, $p < .001$. The interaction between face race and participant group was also significant, $F(1, 32) = 5.19$, $MSE = .02$, $p < .03$. The three-way interaction between face race, orientation, and participant group was also significant, $F(1, 32) = 4.23$, $MSE = .13$, $p < .048$.

To clarify the source of the interaction, separate ANOVAs were conducted for European and Asian participants.

3.2.1.1. European participants. The ANOVA revealed significant main effects of face race, $F(1, 17) = 29.04$, $MSE = .09$, $p < .001$, and orientation, $F(1, 17) = 27.90$, $MSE = .111$, $p < .001$. Both the inversion effect and own-race advantage were demonstrated.

3.2.1.2. Asian participants. Only the main effect of orientation was significant, $F(1, 15) = 31.073$, $MSE = .09$, $p < .001$.

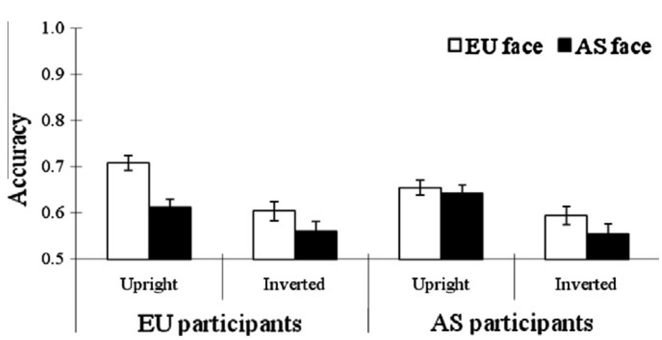

Fig. 6. Means and standard errors of accuracy in different conditions of Experiment 2.

3.2.2. Chance level t-tests

The performances in all the eight conditions were significantly higher than chance level ($ps < .012$).

3.3. Discussion

The results show that participants were able to match a frontal-view image to its corresponding one-tone silhouette image even though all the view-based information, such as the facial texture and pictorial matching, was not available.

The principal patterns of Experiment 1 remain unchanged in Experiment 2. The face inversion effect and the asymmetric race effect obtained in Experiment 1 were replicated in Experiment 2. These results suggest that the 3D face representations and the underlying processes might be restricted to some specific face expertise, e.g., humans are only experts in recognizing upright faces and own-race faces. The asymmetrical race effect reveals that European participants still performed better in the recognition of European faces than Asian faces, while Asian participants performed equally well between the two different ethnic faces.

4. Experiment 3

In the previous two experiments, the 3D information in the faces was essential during the tasks. Experiment 3 serves as a control experiment to examine how participants perform when the 3D information in faces is hard to retrieve. The frontal-view faces were replaced by line-drawing faces. The line-drawing face images which are free from any shadow and shading information contain hardly any cues for 3D representation (Bruce & Langton, 1994; Kemp, Pike, White, & Musselman, 1996; Liu, Collin, Burton, & Chaudhuri, 1999) and thus are more difficult to recognize than normal faces (Davies, Ellis, & Shepherd, 1978; Leder, 1996; Rhodes, Brennan, & Carey, 1987). The lack of sufficient 3D information cues in line-drawing faces might force participants to rely only on other information to recognize the silhouettes, e.g., the estimation of the size of the facial feature (nose, mouth, chin, etc.) or the distance between facial features. This experiment can further examine the contribution of these strategies.

4.1. Method

4.1.1. Participants

Eighteen European students (3 male, mean age = 25.52 years) from Switzerland and eighteen students (8 male, mean age = 20.81 years) from Taiwan participated in Experiment 3.

4.1.2. Materials

The face images were identical to those used in Experiments 1 and 2. However, the targeting frontal-view faces were converted to line-drawing faces. The line drawings of the faces were drawn manually from the original faces using Photoimpact 10. After the

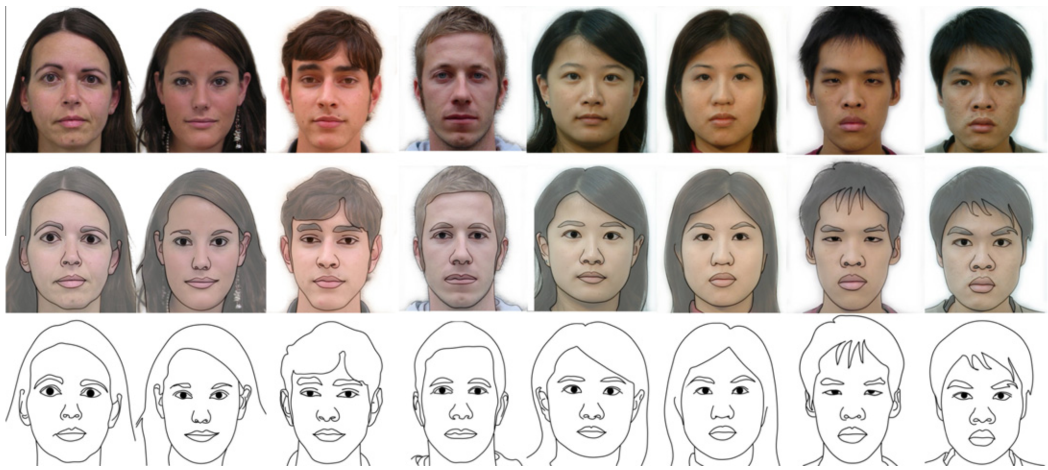

Fig. 7. Examples of the front-view faces with their corresponding line-drawing faces in Experiment 3.

line drawings of the faces were completed, the original face was deleted and only the line-drawing faces were left. All the line-drawing faces were created in the same manner by the same experimenter. Fig. 7 displays examples of the front-view faces with their corresponding line-drawing faces. The faces in the middle row display the superimposition of the two types of face images.

4.1.3. Procedure

The design and procedure were identical to those in Experiment 2. However, the priming front-view face images were replaced by line-drawing faces. Fig. 8 illustrates the procedure and the time course of the stimulus presentation in Experiment 3.

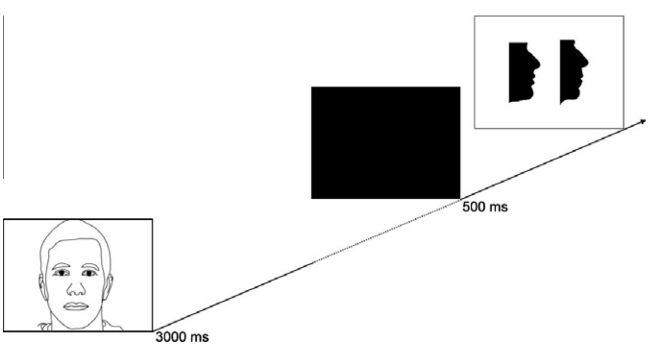

Fig. 8. The procedure and the time course of the stimulus presentation in Experiment 3.

4.2. Results

4.2.1. Accuracy

Fig. 9 displays means and standard errors of the accuracy in Experiment 3. The repeated measures ANOVA revealed only the main effect of orientation, $F(1, 34) = 12.70$, $MSE = .04$, $p < .001$.

4.2.2. Chance level t-tests

The performance was equal to chance level in the recognition of inverted Asian faces for Asian participants ($p > .396$). Performances in all the other conditions were significantly higher than chance level ($ps < .034$), although the accuracy is very low.

4.3. Discussion

The results of Experiment 3 show that the performances were still higher than chance level in most of the conditions. This

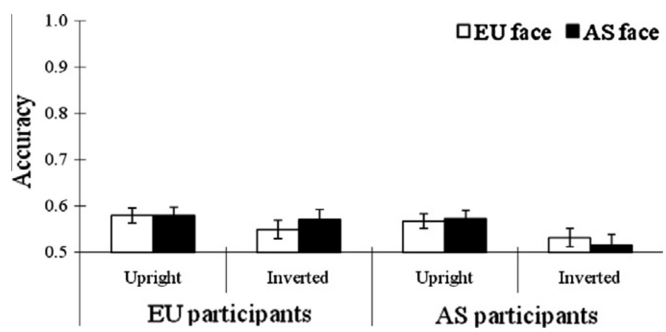

Fig. 9. Means and standard errors of accuracy in different conditions of Experiment 3.

suggests that participants might still rely on some information contained in the line drawings to accomplish the task. One plausible candidate might be the estimation of the sizes or shape of the mouth, nose, or chin, especially when the differences of the size are substantial between the two testing faces, as shown in Fig. 10. In these examples, the sizes of the mouths between the two testing faces differ substantially, and this can easily be observed. These matching strategies might explain why the accuracy remains higher than chance level in most of the conditions when the 3D information in faces is eliminated.

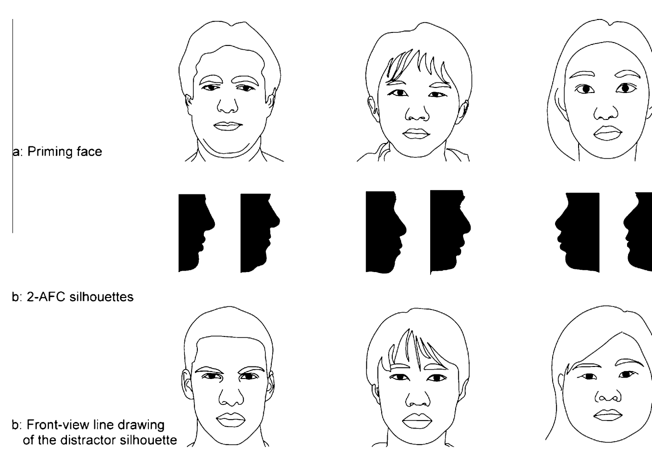

Fig. 10. Illustrations of the substantial size difference of the mouths between the target and the distractor faces.

Nevertheless, the contribution from these matching strategies is relatively low; the performance significantly declined when using line-drawing faces, in which the essential 3D information is substantially reduced (Davies et al., 1978; Rhodes et al., 1987). The lack of 3D information in faces might be critical in explaining this low performance, as it has been found that the addition of shading information can significantly improve the recognition of cartoon (line-drawing) faces (Bruce, Hanna, Dench, Healey, & Burton, 1992). The influence of the insufficient 3D information might lead to an even more severe impact on the performance when the task directly examine the 3D representations as in current experiments.

The results of Experiment 3 also strengthen the application of 3D representation in Experiment 2. After all, participants did not have a depth-rotated representation of the frontal-view face in their memory, because the faces were unfamiliar to them in Experiment 2. Moreover, as described in the introduction section, multiple 2D view-based algorithms are not applicable under the circumstances of these experiments (Poggio & Edelman, 1990; Tarr & Pinker, 1989; Ullman & Basri, 1991). Various 2D view-based theories of face recognition seem insufficient to explain the results that participants can readily match the one-tone black silhouettes to the corresponding frontal-view faces. Furthermore, as Experiment 3 revealed, matching the silhouette based on the sizes or distance between features was inefficient.

5. Comparisons between Experiments 1–3

5.1. Pooled accuracy & size of inversion effect

Only the formats of face stimuli were altered between Experiments 1–3. In Experiment 1, participants could match the front-view faces to the profile faces not only by the 3D representations but also by a range of pictorial information, such as the texture information, the size of features, and the distances between features. However, in Experiment 2, the 3D representations of the faces were essential information for the accomplishment of the task of matching the front-view faces to the silhouette images. Any 2D view-based representation was apparently inapplicable. In Experiment 3, using line-drawing images, the 3D representations of faces are hardly constructable, and participants might then rely on different strategies to complete the task, for instance estimation of the size or shape of the facial features, or the relative distances between the face features. The comparisons between different experiments may provide important information regarding the characteristics of face representation in face images with different formats. To compare the performance in different experiments, all the data were pooled together by creating an extra factor, i.e. experiment (three levels: Experiment 1, Experiment 2, and Experiment 3). A repeated measures analysis of variance (ANOVA) was conducted with face race and orientation as within-participant factors, and participant group and experiment as between-participant factors. Only the main effects and the interaction effects concerning the new factor of experiment were reported.

The repeated measures ANOVA revealed the main effects of experiment, $F(2, 102) = 199.68$, $MSE = 2.66$, $p < .001$. Bonferroni-adjusted posteriori pairwise comparison revealed that all the pairwise comparisons between experiments were significant. Performance was better in Experiment 1 ($M = .81$) than Experiment 2 ($M = .62$), and performances in both Experiments 1 and 2 were better than that in Experiment 3 ($M = .59$). The interaction between experiment and orientation, $F(2, 102) = 16.76$, $MSE = .06$, $p < .01$, and between experiment and face race were also significant, $F(2, 102) = 11.07$, $MSE = .04$, $p < .01$. The three-way interaction between experiment, face race, and participant group was also significant, $F(2, 102) = 7.7$, $MSE = .03$, $p < .01$.

The comparison analyses between the experiments reveal an interaction effect between experiment and orientation. However, as all the three experiments reveal significant inversion effects, we conducted separate ANOVAs to examine the fluctuation of the size of inversion effect between experiments. The difference in accuracy between upright and inverted trials for each participant was calculated so as to quantify the size of inversion effect. The data were used to conduct a three-way ANOVA with face race as within-participant factor, and participant group and experiment as between-participant factors. The repeated measures ANOVA revealed the main effects of Experiment, $F(2, 102) = 16.76$, $MSE = .12$, $p < .001$. Bonferroni-adjusted posteriori pairwise comparison revealed that all the pairwise comparisons between experiments were significant. It shows that the size of inversion effect is larger in Experiment 1 ($M = .12$) than Experiment 2 ($M = .08$), and both of them are larger than that in Experiment 3 ($M = .03$). The interaction between face race and participant group was also significant, $F(1, 102) = 7.1$, $MSE = .03$, $p < .01$, as illustrated in Fig. 11. It shows that participants seem to have a larger inversion effect for the faces of their own ethnic group. However, simple main effects reveal that the size of the inversion effect was comparable for both participant groups on European faces, whereas it was larger for the Asian participant group than the European participant group on Asian faces.

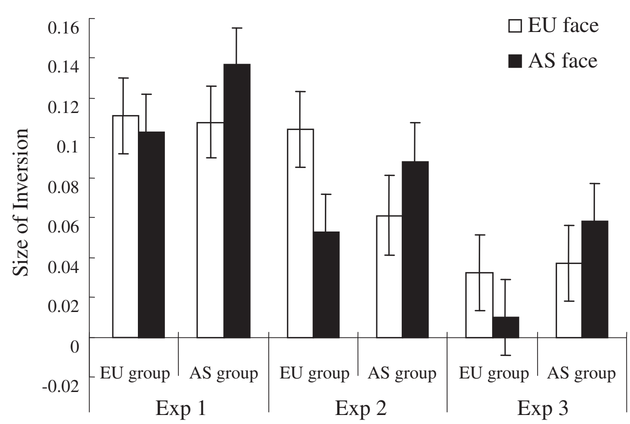

Fig. 11. Interaction between face race and participants group on the size of inversion effect in different experiments.

5.2. Gender difference

Across the three experiments, there were some considerable differences in the gender composition between the Asian and European participant populations. In sum there were 45 European female, 9 European male, 30 Asian female, and 24 Asian male participants in the experiments. The European male participants were far fewer than the Asian male participants, whereas European female participants substantially outnumbered Asian female participants. It is unclear whether the gender of participants might mediate with other factors manipulated in the experiments. To examine the role of participant gender, we pooled the data across the three experiments with an extra factor of participant gender. A five-way ANOVA with orientation and face race as within-participant factors, and experiment, participant gender, and participant group as between-participant factor was conducted. The repeated measures ANOVA revealed main effects of Exp, $F(2, 96) = 157.626$, $MSE = 2.15$, $p < .001$, face race, $F(1, 96) = 23.909$, $MSE = .086$, $p < .001$, and orientation, $F(1, 96) = 90.153$, $MSE = .336$, $p < .001$. The interactions between orientation and Exp, $F(2, 96) = 11.838$, $MSE = .044$, $p < .001$; between face race and participant group, $F(1, 96) = 7.76$, $MSE = .028$, $p < .006$; and between face race and

Exp, $F(2, 96) = 8.184$, $MSE = .029$, $p < .001$, were also significant. The three-way interactions between face race, participant group, and Exp, $F(2, 96) = 5.703$, $MSE = .02$, $p < .005$, and between face race, participant group, and orientation, $F(1, 96) = 6.123$, $MSE = .016$, $p < .015$, were also significant. No other main effects or interaction effects were found. Note that the factor of participant gender revealed neither a main effect, $F(1, 96) = 2.207$, $MSE = .038$, $p = .141$, nor any interactions with other factors ($ps > .14$). It suggests that the factor of participant gender does not lead to any main effect or interaction with any other factors in this study.

6. General discussion

In this paper, we adopted a new paradigm to investigate the application and characteristics of 3D face representations. Participants were required to match frontal views of faces to their profiles (Experiment 1) or silhouettes (Experiment 2), or to match line-drawing faces to their silhouettes (Experiment 3). The results reveal that participants can readily solve the tasks when the face images retain the information essential for the formation of a 3D face representation. The performance substantially declined when the 3D information in faces was eliminated.

These results not only provide converging evidence to support the existence of 3D face representations, but also reveal important characteristics of such representations in different ethnic groups and in different formats of face images. In Experiments 1 and 2, only the frontal-view face was presented. Accordingly, various view-based transformation algorithms are not applicable to generate the profile (or silhouette) view from the frontal-view face due to the singularity of the frontal-view face (Schyns & Bulthoff, 1993). The results of Experiment 2 provide evidence supporting the application of 3D face representations. The possibility that participants might base their judgment on the size or shape of face features, or inter-distance between features further was further examined in Experiment 3. In addition, different behavioral patterns were demonstrated when participants were required to recognize faces in different formats. In Experiments 1 and 2, participants' performances were higher than chance level in all different conditions when plentiful 3D information is accessible. However, the performances severely declined in Experiment 3, in which 3D information is substantially reduced.

Another interesting finding in the research is the fluctuation of the inversion effects in different experiments. The inversion effect was demonstrated when the tasks demanded explicit application of 3D face representations, whereas it was substantially weakened when the 3D information in faces was eliminated (Experiment 3). The emergence of an inversion effect in Experiment 2 provides especially important evidence to contradict the assumption that participants in the experiment merely rely on the distances between features to match the faces. Schwaninger, Ryf, and Hofer (2003) found that the inversion effect disappeared when participants were required to compare the distances between face features during face recognition tasks. Merely comparing the distances between face features is a relatively lower level processing which would not lead to the emergence of an inversion effect. The results of Experiment 3 correspond with this inference.

Although the existence of the 3D face representations is supported in the current research, the underlying mechanisms of 3D face representations seem different from the representations defined by current 3D representational models, such as object-centered models (Biederman, 1987; Marr, 1982) or the 3D computational approach to face recognition by Blanz and Vetter (1999). The major difference is that the 3D representations of faces may not be a purely viewpoint-independent 3D face representation, but a representation based on expertise that favors the transformation in an upright orientation and in an own-race face. We assume that a coarse prototype 3D face model that favors the upright transformation is built into the human brain through exposure to numerous upright and different viewpoint faces in daily lives. Meanwhile, high frequency of exposure to a specific face might help us to sculpture the coarse prototype 3D face into a finer one. Finally, with the application of the 3D models in the recognition of numerous faces, the 3D face model seems to become applicable also to unfamiliar faces.

References

Alley, T. R., & Schultheis, J. A. (2001). Is facial skin tone sufficient to produce a cross-racial identification effect? *Perceptual and Motor Skills, 92*, 1191–1198.

Barden, J., Maddux, W. W., Petty, R. E., & Brewer, M. B. (2004). Contextual moderation of racial bias: The impact of social roles on controlled and automatically activated attitudes. *Journal of Personality & Social Psychology, 87*, 5–22.

Beymer, D., & Poggio, T. (1995). Face recognition from one example view. In *Fifth international conference on computer vision (ICCV'95)* (pp. 500–507).

Biederman, I. (1987). Recognition-by–components: A theory of human image understanding. *Psychological Review, 94*, 115–147.

Biederman, I. (2000). Recognizing depth-rotated objects: A review of recent research and theory. *Spatial Vision, 13*(2, 3), 241–253.

Biederman, I., & Gerhardstein, P. C. (1993). Recognizing depth-rotated objects – Evidence and conditions for 3dimensional viewpoint invariance. *Journal of Experimental Psychology–Human Perception and Performance, 19*, 1162–1182.

Biederman, I., & Kalocsai, P. (1997). Neurocomputational bases of object and face recognition. *Philosophical Transactions of the Royal Society of London Series B–Biological Sciences, 352*(1358), 1203–1219.

Blanz, V., & Vetter, T. (1999). A morphable model for the synthesis of 3D faces. In *Proceedings of the 26th annual conference on computer graphics and interactive techniques* (pp. 187–194).

Bruce, V. (1988). *Recognising faces*. London: Lawrence Erlbaum.

Bruce, V., Hanna, E., Dench, N., Healey, P., & Burton, M. (1992). The importance of 'Mass' in line drawings of faces. *Applied Cognitive Psychology, 6*, 619–628.

Bruce, V., Healey, P., Burton, M., Doyle, T., Coombes, A., & Linney, A. (1991). Recognition facial surfaces. *Perception, 20*, 755–769.

Bruce, V., & Langton, S. (1994). The use of pigmentation and shading information in recognizing the sex and identities of faces. *Perception, 23*, 803–822.

Bülthoff, H. H., & Edelman, S. (1992). Psychophysical support for a 2-dimensional view interpolation theory of object recognition. *Proceedings of the National Academy of Sciences of the United States of America, 89*, 60–64.

Bülthoff, H. H., Edelman, S. Y., & Tarr, M. J. (1995). How are three-dimensional objects represented in the brain? *Cerebral Cortex, 5*, 247–260.

Casey, S. J., & Newell, F. (2003). *Haptic own-face recognition. Poster presented at EuroHaptics 2003*, Zurich.

Chen, C. C., Kao, K. L. C., & Tyler, C. W. (2006). Face configuration processing in the human brain: The role of symmetry. *Cerebral Cortex, 17*, 1423–1432.

Davidenko, N. (2007). Silhouetted face profiles: A new methodology for face perception research. *Journal of Vision, 7*(4), 1–17.

Davies, G. M., Ellis, H. D., & Shepherd, J. W. (1978). Face recognition accuracy as a function of mode of representation. *Journal of Applied Psychology, 63*, 180–187.

Desimone, R. (1991). Face-selective cells in the temporal vortex of monkeys. *Journal of Cognitive Neuroscience, 3*, 1–7.

Ellis, H. D., & Florence, M. (1990). Bodamer's (1947) paper on prosopagnosia. *Cognitive Neuropsychology, 7*, 81–105.

Farah, M. J. (1991). Patterns of co-occurrence among the associative agnosias: Implications for visual object representation. *Cognitive Neuropsychology, 8*, 1–19.

Farah, M. J., Levinson, K. L., & Klein, K. L. (1995). Face perception and within-category discrimination in prosopagnosia. *Neuropsychologia, 33*, 661–674.

Goldstein, A. G., & Chance, J. (1979). Do 'foreign' faces really look alike? *Bulletin of the Psychonomic Society, 13*(2), 111–113.

Grill-Spector, K., Kushnir, T., Avidan, G., Edelman, S., Itzchak, Y., & Malach, R. (1999). Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron, 24*, 187–203.

Hayward, W. G., Rhodes, G., & Schwaninger, A. (2008). An own-race advantage for components as well as configurations in face recognition. *Cognition, 106*, 1017–1027.

Hecaen, H., & Angelergues, R. (1962). Agnosia for faces (prosopagnosia). *Archives of Neurology, 7*, 92–100.

Hill, H., Schyns, P. G., & Akamatsu, S. (1997). Information and viewpoint dependence in face recognition. *Cognition, 62*, 201–222.

Johnson, A., Hill, H., & Carman, N. (1992). Recognising faces: Effects of lighting direction, inversion, and brightness reversal. *Perception, 21*, 365–375.

Kanwisher, N., Downing, P., Epstein, R., & Kourtzi, Z. (2001). Functional imagining of visual recognition. In R. Cabeza & A. Kingstone (Eds.), *Handbook of functional neuroimaging of cognition* (pp. 109–152). Cambridge, MA: MIT Press.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience, 17*, 4302–4311.

Kemp, R., Pike, G., White, P., & Musselman, A. (1996). Perception and recognition of normal and negative faces: The role of shape from shading and pigmentation cues. *Perception, 25*, 37–52.

Kilgour, A. R., & Lederman, S. J. (2002). Face recognition by hand. *Perception & Psychophysics, 64*, 339–352.

Lades, M., Vorbrüggen, J. C., Buhmann, J., Lange, J., von der Malsburg, C., Würtz, R. P., et al. (1993). Distortion invariant object recognition in the dynamic link architecture. *IEEE Transactions on Computers, 42*(3), 300–311.

Leder, H. (1996). Line drawings of faces reduce configural processing. *Perception, 25*, 355–366.

Liu, C. H., Collin, C. A., Burton, A. K., & Chaudhuri, A. (1999). Lighting direction affects recognition of untextured faces in photographic positive and negative. *Vision Research, 39*(24), 4003–4009.

Marr, D. (1982). *Vision*. San Franciso: W.H. Freeman.

Marr, D., & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society, London, B, 200*, 269–294.

Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law, 7*, 3–35.

ÓToole, A. J., Bülthoff, H. H., Troje, N. F., &Vetter, T. (1995). Face recognition across large view point changes. In M. Bichsel (Ed.), *Proceedings of the intern workshop on automatic face- and gesture recognition* (pp. 326–331). Zürich: MultiMedia Lab. Computer Science Dept. Univ.

Ojemann, J. G., Ojemann, G. A., & Lettich, E. (1992). Neuronal activity related to faces and matching human right nondominant temporal cortex. *Brain, 115*, 1–13.

Perrett, D. I., Oram, M. W., Harries, M. H., Bevan, R., Hietanen, J. K., Benson, P. J., et al. (1991). Viewer-centred and object-centred coding of heads in the macaque temporal cortex. *Experimental Brain Research, 86*, 159–173.

Poggio, T. (1990). 3D object recognition: On a result by Basri and Ullman. Technical report 9005-03, IRST, Povo, Italy.

Poggio, T. (1991). 3D object recognition and prototypes: One 2D view may be sufficient. Technical Report 9107-02, IRST, Povo, Italy.

Poggio, T., & Vetter, T. (1992). *Recognition and structure from one 2D model view: Observations on prototypes, object classes and symmetries*. A. I. Memo No. 1347, Artificial Intelligence Laboratory, Massachusettes Institute of Technology.

Poggio, T., & Edelman, S. (1990). A network that learns to recognize 3dimensional objects. *Nature, 343*(6255), 263–266.

Rhodes, G., Brennan, S. E., & Carey, S. (1987). Identification and ratings of caricatures: Implications for mental representations of faces. *Cognitive Psychology, 19*, 473–497.

Russell, R., Biederman, I., Nederhouser, M., & Sinha, P. (2007). The utility of surface reflectance for the recognition of upright and inverted faces. *Vision Research, 47*, 157–165.

Schwaninger, A., Lobmaier, J., Wallraven, C., & Collishaw, S. (2009). Two routes to face perception: Evidence from psychophysics and computational modeling. *Cognitive Science, 33*, 1413–1440.

Schwaninger, A., Schumacher, S., Wallraven, C., & Bülthoff, H. H. (2007). Using 3D computer graphics for perception: The role of local and global information in face processing. In *Proceedings of the 4th symposium on applied perception in graphics and visualization* (pp. 19–26). New York, USA: ACM Press.

Schwaninger, A., Ryf, S., & Hofer, F. (2003). Configural information is processed differently in perception and recognition of faces. *Vision Research, 43*, 1501–1505.

Schwaninger, A., Wallraven, C., & Bülthoff, H. H. (2004). Computational modeling of face recognition based on psychophysical experiments. *European Journal of Psychology, 63*, 207–215.

Schwaninger, A., Wallraven, C., Cunningham, D. W., & Chiller-Glaus, S. (2006). Processing of identity and emotion in faces: A psychophysical, physiological and computational perspective. *Progress in Brain Research, 156*, 321–343.

Schyns, P. G., & Bulthoff, H. H. (1993). *Conditions for viewpoint dependent face recognition*. A.I. Memo No. 1432, Artificial Intelligence Laboratory, Massachusetts Institute of Technology.

Tanaka, J. W., & Farah, M. J. (1991). Second-order relational properties and the inversion effect: Testing a theory of face perception. *Perception & Psychophysics, 50*, 367–372.

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *The Quarterly Journal of Experimental Psychology, 46A*(2), 225–245.

Tanaka, J. W., Kiefer, M., & Bukach, C. M. (2004). A holistic account of the own-face effect in face recognition: Evidence from a cross-cultural study. *Cognition, 93*, B1–B9.

Tarr, M. J., & Bülthoff, H. H. (1995). Is human object recognition better described by Geon structural descriptions or by multiple views – Comment on Biederman and Gerhardstein (1993). *Journal of Experimental Psychology–Human Perception and Performance, 21*, 1494–1505.

Tarr, M. J., & Pinker, S. (1989). Mental rotation and orientation-dependence in shape-recognition. *Cognitive Psychology, 21*, 233–282.

Tong, F., & Nakayama, K. (1999). Robust representations for faces evidence from visual search. *Journal of Experimental Psychology Human Perception and Performance, 25*, 1016–1035.

Troje, N. F., & Kersten, D. (1999). Viewpoint-dependent recognition of familiar faces. *Perception, 28*, 483–487.

Ullman, S., & Basri, R. (1991). Recognition by linear-combinations of models. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 13*, 992–1006.

Valentine, T. (1988). Upside-down faces: A review of the effect of inversion upon face recognition. *British Journal of Psychology, 79*, 471–491.

Vetter, T. (1998). Synthesis of novel views from a single face image. *International Journal of Computer Vision, 28*, 103–116.

Wallraven, C., Schwaninger, A., & Bülthoff, H. H. (2005). Learning from humans: Computational modeling of face recognition. *Network: Computation in Neural Systems, 16*, 401–418.

Wallraven, C., Schwaninger, A., Schuhmacher, S., & Bülthoff, H. H. (2002). View-based recognition of faces in man and machine: Re-visiting inter–extra–ortho. *Lectures Notes in Computer Science, 2525*, 651–660.

Wiskott, L., Fellous, J. M., Krüger, N., & von der Malsburg, C. (1997). Face recognition by elastic bunch graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 19*, 775–779.

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology, 81*, 141–145.

Yin, R. K. (1970). Face recognition by brain-injured patients: A dissociable ability? *Neuropsychologia, 8*, 395–402.

Yip, A. W., & Sinha, P. (2002). Contribution of color to face recognition. *Perception, 32*, 995–1003.